Štěpán Papáček; Jiří Jablonský; Ctirad Matonoha

On two methods for the parameter estimation problem with spatio-temporal FRAP data

In: Jan Chleboun and Petr Přikryl and Karel Segeth and Jakub Šístek and Tomáš Vejchodský (eds.): Programs and Algorithms of Numerical Mathematics, Proceedings of Seminar. Dolní Maxov, June 8-13, 2014. Institute of Mathematics AS CR, Prague, 2015. pp. 163–168.

Persistent URL: http://dml.cz/dmlcz/702679

Programs and Algorithms of Numerical Matematics 17
J. Chleboun, P. Přikryl, K. Segeth, J. Šístek, T. Vejchodský (Eds.)
Institute of Mathematics AS CR, Prague 2015

ON TWO METHODS FOR THE PARAMETER ESTIMATION PROBLEM WITH SPATIO-TEMPORAL FRAP DATA

Štěpán Papáček[1], Jiří Jablonský[1], Ctirad Matonoha[2]

[1] University of South Bohemia in České Budějovice,
Faculty of Fisheries and Protection of Waters, CENAKVA, Institute of Complex Systems
Zámek 136, 373 33 Nové Hrady, Czech Republic
spapacek@frov.jcu.cz, jiri.jablonsky@gmail.com
[2] Institute of Computer Science, Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2, 182 07 Prague 8, Czech Republic
matonoha@cs.cas.cz

Abstract

FRAP (Fluorescence Recovery After Photobleaching) is a measurement technique for determination of the mobility of fluorescent molecules (presumably due to the diffusion process) within the living cells. While the experimental setup and protocol are usually fixed, the method used for the model parameter estimation, i.e. the data processing step, is not well established. In order to enhance the quantitative analysis of experimental (noisy) FRAP data, we firstly formulate the inverse problem of model parameter estimation and then we focus on how the different methods of data pre-processing influence the confidence interval of the estimated parameters, namely the diffusion constant p. Finally, we present a preliminary study of two methods for the computation of a least-squares estimate $\hat{p}$ and its confidence interval.

1. Introduction

The FRAP technique is based on measuring the change in fluorescence intensity in a region of interest (ROI - generally a Euclidean 2D or 3D domain). These changes are induced by an external stimulus, a high-intensity laser pulse provided by the CLSM (Confocal Laser Scanning Microscopy). The stimulus, also called *bleaching*, causes an (ir)reversible loss in fluorescence in the bleached area without any damage to intracellular structures. After the bleach, the observed recovery in fluorescence reflects the mobility (related to diffusion) of fluorescent compounds from the area outside the bleach.

Based on spatio-temporal 2D FRAP images, the process of diffusive transport can be reconstructed using either a closed form model or a numerical simulation based model. In this paper, we study both approaches. We show the results for the oversimplified one-spatial-point Moullineaux method [4] and the results based on the numerical integration of the Fick diffusion PDE (Partial Differential Equation) with the realistic initial and boundary conditions [5].

2. Parameter estimation based on spatio-temporal data

We aim to present a parameter estimation problem with spatio-temporal experimental observation in a comprehensive mathematical framework allowing simultaneously to determine both the parameter value p (generally $p \in \mathbb{R}^q,\ q \in \mathbb{N}$)[1] and the corresponding confidence interval proportional to the output noise and a quantity related to the sensitivity, see (7). The data are represented by a (measured) signal on a Cartesian product of the space-points $(x_i)_{i=1}^n$ and time-points $(t_j)_{j=1}^m$; let $N_{\text{Data}} := m \times n$ be the total number of spatio-temporal data points. We define the operator $S : \mathbb{R}^q \to \mathbb{R}^{N_{\text{Data}}}$ that maps parameter values $p_1, \ldots, p_q$ to the solution of the underlying initial-boundary value problem, e.g. (9), evaluated at points (x_i, t_j):

$$S(p) = \{y(x_i, t_j, p) \in \mathbb{R}, \quad 1 \le i \le n, \quad 1 \le j \le m\}. \tag{1}$$

Some commonly used FRAP methods do not employ all the N_{Data} measured values at points $\{(x_i, t_j),\ i = 1, \ldots, n,\ j = 1, \ldots, m\}$. They either employ some of the values or perform some preprocessing (e.g. space averaging, see [6]). Hence we further define the observation operator $G : \mathbb{R}^{N_{\text{Data}}} \to \mathbb{R}^{N_{\text{data}}}$ that evaluates the set of values $S(p)$ on a certain subset of the full data space ($N_{\text{data}} \le N_{\text{Data}}$):

$$G(S(p)) = (z(x_l, t_l, p))_{l=1}^{N_{\text{data}}} \tag{2}$$

We now define the forward map $F : p \to z(x_l, t_l, p)_{l=1}^{N_{\text{data}}}$. Here, $F = G \circ S$ represents the parameter-to-output map, defined as the composition of the PDE solution operator S and the observation operator G.[2] Our regression model is now

$$F(p) = \text{data}, \tag{3}$$

where the data are modeled as contaminated with additive Guassian noise

$$\text{data} = F(p_T) + e = (z(x_l, t_l, p_T))_{l=1}^{N_{\text{data}}} + (e_l)_{l=1}^{N_{\text{data}}}.$$

Here $p_T \in \mathbb{R}^q$ denotes the true values and $e \in \mathbb{R}^{N_{\text{data}}}$ is a data error vector which we assume to be normally distributed with variance σ^2, i.e. $e_i = \mathcal{N}(0, \sigma^2)$ $i = 1, \ldots, N_{\text{data}}$.

Given some data, the aim of the parameter estimation problem is to find p_T, such that (3) is satisfied in some appropriate sense. Since (3) usually consists of an overdetermined system (there are more data points than unknowns), it cannot be expected that (3) holds with equality, but instead an appropriate notion of a solution is that of a least-squares solution $\hat{p}$ (with $\| \ . \ \|$ denoting the Euclidean norm on $\mathbb{R}^{N_{\text{data}}}$):

$$\| \text{ data} - F(\hat{p}) \|^2 = \min_p \| \text{ data} - F(p) \|^2 . \tag{4}$$

[1] We prefer this more general definition of the model parameter vector instead of the single scalar quantity because we aim to work with more complex model than (9) in the near future.

[2] For the one-point Moullineaux method [4], only the point with the spatial coordinate $x = 0$ is measured, i.e. $G_M : z(t_j, p) := z(0, t_j, p) = y(0, t_j, p),\ j = 1, \ldots, N_{\text{data}} = m$. For the second method, we reduce the data space taking the so-called relevant data only [6], i.e. $G_{\text{PDE}} : z(x_l, t_l, p) = y(x_i, t_j, p),\ i = 1, \ldots, n^* \le n,\ j = 1, \ldots, m^* \le m,\ l = 1, \ldots, N_{\text{data}} = m^* \times n^*$.

Sensitivity analysis and confidence intervals

For the sensitivity analysis we require the Fréchet-derivative $F'[p_1, \dots, p_q] \in \mathbb{R}^{N_{\text{data}} \times q}$ of the forward map F, that is

$$\begin{aligned} F'[p_1, \dots, p_q] &= \begin{pmatrix} \frac{\partial}{\partial p_1} F(p_1, \dots, p_q) & \dots & \frac{\partial}{\partial p_q} F(p_1, \dots, p) \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial}{\partial p_1} z(x_1, t_1, p) & \dots & \frac{\partial}{\partial p_q} z(x_1, t_1, p) \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ \frac{\partial}{\partial p_1} z(x_{N_{\text{data}}}, t_{N_{\text{data}}}, p) & \dots & \frac{\partial}{\partial p_q} z(x_{N_{\text{data}}}, t_{N_{\text{data}}}, p) \end{pmatrix}. \end{aligned}$$

A corresponding quantity is the Fisher information matrix (FIM)

$$M[p_1, \dots, p_q] = F'[p_1, \dots, p_q]^T F'[p_1, \dots, p_q] \in \mathbb{R}^{q \times q}. \tag{5}$$

Based on the book of Bates and Watts [1], we can estimate confidence intervals. Suppose we have computed $\hat{p}$ as a least-squares solution in the sense of (4). Let us define the residual as

$$res^2(\hat{p}) = \|F(\hat{p}) - \text{data}\|^2 = \sum_{i=1}^{N_{\text{data}}} \left[\text{data}_i - z(x_i, t_i, \hat{p})\right]^2. \tag{6}$$

Then according to [1], it is possible to quantify the error between the computed parameters $\hat{p}$ and the true parameters p_T.

Having only one single scalar parameter p as unknown, the Fisher information matrix M collapses into the scalar quantity $\sum_{i=1}^{N_{\text{data}}} \left[\frac{\partial}{\partial p} z(x_i, t_i, p) \mid_{p=\hat{p}}\right]^2$, and the $1-\alpha$ confidence interval for full observations is described as follows

$$(\hat{p} - p_T)^2 \sum_{i=1}^{N_{\text{data}}} \left[\frac{\partial}{\partial p} z(x_i, t_i, p) \mid_{p=\hat{p}}\right]^2 \leq \frac{res^2(\hat{p})}{N_{\text{data}} - 1} f_{1, N_{\text{data}}-1}(\alpha), \tag{7}$$

where $f_{1, N_{\text{data}}-1}(\alpha)$ corresponds to the upper α quantile of the Fisher distribution with 1 and $N_{\text{data}} - 1$ degrees of freedom.

In (7), several simplifications are possible. Note that according to our noise model, the residual term $\frac{res^2(\hat{p})}{N_{\text{data}}-1}$ is an estimator of the error variance [1] such that the approximation

$$\frac{res^2(\hat{p})}{N_{\text{data}} - 1} \sim \sigma^2 \tag{8}$$

holds if N_{data} is large. Moreover, we remind the reader that the Fisher distribution with 1 and $N_{\text{data}} - 1$ degrees of freedom converges to the χ^2-distribution as $N_{\text{data}} \to \infty$. Hence, the term $f_{1, N_{\text{data}}-1}(\alpha)$ can approximately be viewed as independent of N_{data} as well and of a moderate size.

3. Two FRAP methods: Assessment of uncertainty

Let us proceed to the FRAP measurement technique [4, 5]. We assume the special geometry residing in one-dimensional simplification getting the measured fluorescent intensity level y as a function of the spatial coordinate x, time t and diffusion coefficient p (generally time dependent, e.g. $p = (p_j)_{j=1}^m$):

$$\frac{\partial y}{\partial t} - p\frac{\partial^2 y}{\partial x^2} = 0 \ , \tag{9}$$

in $(t_0, T)\times\Omega$, with suitable boundary conditions on $(t_0, T)\times\partial\Omega$ and initial conditions in Ω, where $\Omega \subset \mathbb{R}$. Problem (9) represents a reliable model of the FRAP process. The corresponding inverse formulation is used in our software CA-FRAP[3] for the processing of the real FRAP data resulting in the solution vector $(\hat{p}_j)_{j=1}^m$. Here, in this paper, the software CA-FRAP is further used (in Subsection 3.2) for the simulation of virtual FRAP data and the subsequent evaluation of the FIM.

According to [2], the standard error of a parameter p_k estimate, i.e. $SE(\hat{p}_k)$, is

$$SE(\hat{p}_k) = \hat{\sigma}\sqrt{{M_{kk}}^{-1}}, \tag{10}$$

where $\hat{\sigma}$ is the data error variance estimate. Relation (10) highlights the importance of the FIM and is further used for the comparison of two FRAP data processing methods.

3.1. The one-point Moullineaux method

C. W. Moulineaux *et al.* [4] measured one-dimensional bleaching profiles (with common variance σ^2) along the specimen long axis. They took the ROI as coincident with the real axis ($x \in \mathbb{R}$) and the initial bleaching profile (of bleached particles) as the Gaussian with half-width r_0 at height $y_{0,0}e^{-2}$, i.e. $y(x, t_0, p) = y_{0,0}\exp\frac{-2x^2}{{r_0}^2}$. Here t_0 corresponds to the initial time and can be set to zero. Then, the solution $y(x, t, p)$ of the diffusion equation (9) for the bleached particles is

$$y(x, t, p) = \frac{y_{0,0}\ r_0}{\sqrt{{r_0}^2 + 8pt}}\exp\frac{-2x^2}{{r_0}^2 + 8pt}, \quad x \in \mathbb{R}, \quad t \in [0, T]. \tag{11}$$

The time evolution of the maximum depth $y(0, t, p)$ is taken by Moullineaux *et al.* as the single observed spatial data point $z_M(t, p)$.[4] It holds $z_M(t, p) = \frac{r_0\ y_{0,0}}{\sqrt{{r_0}^2+8pt}}$.

The FIM, based on the semi-relative sensitivities, collapses to a scalar quantity $M_M = \sum_{j=1}^m \left[\frac{\partial z_M(t_j,p)}{\partial p}p\right]^2 = \sum_{j=1}^m \frac{(4r_0pt_j)^2}{({r_0}^2+8pt_j)^3} = \frac{1}{4}\sum_{j=1}^m \frac{(8s_j)^2}{(1+8s_j)^3}$, where $s_j := \frac{pt_j}{{r_0}^2}$ and an estimate $\hat{p}$ is taken instead of p.

[3] See [3, 5] for more details or mail to: matonoha@cs.cas.cz.

[4] The authors of [4] used the weighted linear regression based on equation $z_M(t, p) = \frac{r_0\ y_{0,0}}{\sqrt{{r_0}^2+8pt}}$ in order to estimate the diffusion coefficient p. They calculated neither the FIM nor the standard error of the parameter p estimate using (10).

Let us assume that we have an equidistant spacing $\Delta s := \frac{s_m - s_1}{m-1}$ such that the sum can be approximated by an integral.[5] Then we get the following expression for the FIM (after some algebraic manipulation assisted by the Mathematica software)

$$M_M \approx \frac{m-1}{32(s_m - s_1)} \left[\ln\left(\frac{1+8s_m}{1+8s_1}\right) - \frac{8(s_m - s_1)(1 + 12(s_1 + s_m) + 128 s_1 s_m)}{(1+8s_m)^2(1+8s_1)^2} \right] + \left[\frac{8{s_1}^2}{(1+8s_1)^3} + \frac{8{s_m}^2}{(1+8s_m)^3} \right]. \quad (12)$$

The expression for M_M is positive, increasing with the number of measurement points, i.e. with $T = t_1 + (m-1)\Delta t$ (for fixed Δt and t_1), and represents the lower bound for the FIM as a scalar quantity (when a scalar p is estimated).[6]

3.2. Initial boundary value problem for PDE (9) and the FIM

As the above approach has several limitations, e.g. cell geometry restriction (infinite domain is required), bleach profile must be gaussian-like, etc., we propose to model the diffusion process by the Fick diffusion equation with realistic initial and boundary conditions instead. Then the parameter estimation problem is formulated as an ordinary least squares problem (4) resulting in the estimate $\hat{p}_{PDE}$. This problem is treated elsewhere [3, 5, 6]. Here, we present the uncertainty assessment based on the numerical evaluation of the FIM (implemented in the CA-FRAP). For each time instant t_j we denote $p_j = \hat{p}_{PDE}$. The CA-FRAP solves the inverse problem (9) and takes the simulated output $y(x_i, t_j, p_j)$, $i = 1, \ldots, n$. Then, according to (5), we obtain the FIM (diagonal in this case) using central differences as

$$M^j_{PDE} = \sum_{i=1}^{n} \left[\frac{\partial y(x_i, t_j, p)}{\partial p} \, |_{p=p_j} \right]^2 \approx \sum_{i=1}^{n} \left[\frac{y(x_i, t_j, p_j + \varepsilon) - y(x_i, t_j, p_j - \varepsilon)}{2\varepsilon} \right]^2 \quad (13)$$

where ε is a small positive number. The corresponding quantity M_{PDE} for the estimation of an overall $\hat{p}_{PDE}$ is the sum $\sum_{j=1}^{m} M^j_{PDE}$, cf. (5).

3.3. Numerical evaluation and comparison of the FIM for both method

We have performed several computations of the FIM for both above mentioned approaches. For a particular case $y_{0,0} = r_0 = p = 1$ and the time step between $m = 10$ measurements equal to 0.1, the evaluation of (12) is straightforward and gives $M_M \approx 0.296$ for $s_1 = 0.1$ and $s_m = 1$. The evaluation of M_{PDE} is more complicated. In order to compare both method, the output $y(x_i, t_j, p_j)$ were computed for the same parameter settings as before by solving the forward problem (9), showing the correspondence with (11), indeed. The numerical evaluation of (13) gives then $M_{PDE} \approx 0.363$. We see that the PDE based method, which uses more spatial points at each time level, gives a lower standard error of the estimated parameter p.

[5] The quantities corresponding to the first t_1 and final T measurement time are $s_1 = pt_1/r_0^2$ and $s_m = pT/r_0^2$, $t_1 < T$, respectively.

[6] The one-point Moullineaux method is the simplest method. Other methods, see e.g. [6] for review, use more data points, thus add more (positive) terms to the FIM.

4. Conclusion

We present two methods for the estimation of the fluorescent compounds mobility from the spatio-temporal FRAP measurement. The first and simplest method is based on the curve fitting to a closed formula and needs some unrealistic or hard-to-accomplish conditions. The second method is based on a numerical approximation of the Fick diffusion PDE with either a scalar or time dependent diffusion coefficient p. Both methods are implemented in our software CA-FRAP, which simultaneously provides the parameter estimate (this is not discussed here, in this paper) and the corresponding standard error (using (10)). This aims to promote the following idea across the FRAP community. The bioprocesses are inherently stochastic, thus the mathematical framework related to the model parameter identification should determine both a parameter mean value and a certain confidence interval, which depends on the output noise and the corresponding sensitivity, cf. (10).

Acknowledgements

This work was supported by the MEYS of the Czech Republic – projects "CENAKVA" (No. CZ.1.05/2.1.00/01.0024), "CENAKVA II" (No. LO1205 under the NPU I program), by Postdok JU CZ.1.07/2.3.00/30.0006, and by the long-term strategic development financing of the Institute of Computer Science (RVO:67985807).

References

[1] Bates, D. M. and Watts, D. G.: *Nonlinear regression analysis: Its applications.* John Wiley & Sons, New York, 1988.

[2] Cintrón-Arias, A., Banks, H. T., Capaldi, A., and Lloyd, A. L.: A sensitivity matrix based methodology for inverse problem formulation. J. Inv. Ill-Posed Problems **17** (2009), 545–564.

[3] Matonoha, C. and Papáček, Š.: On the connection and equivalence of two methods for solving an ill-posed inverse problem based on FRAP data. J. Comput. Appl. Math., submitted.

[4] Moullineaux, C. W., Tobin, M. J., and Jones, G. R.: Mobility of photosynthetic complexes in thylakoid membranes. Nature **390** (1997), 421–424.

[5] Papáček, Š., Kaňa, R., and Matonoha, C.: Estimation of diffusivity of phycobilisomes on thylakoid membrane based on spatio-temporal FRAP images. Math. Comput. Modelling **57** (2013), 1907–1912.

[6] Papáček, Š., Jablonský, J., Kaňa, R., Matonoha, C., and Kindermann, S.: From data processing to experimental design and back again: A parameter identification problem based on FRAP images. Accepted for publication in ICDIP 2015: XIII International Conference on Digital Image Processing, Dubai 2015.